Kari Korpinen 012218686

Problem 4

a)
movielens download and extracting. “Data” folder size is 384 Gb. In folder exists loadmovieslens.py file, that works like input reader. Adding ASSG2_TMP.py template python-file to same folder with loadmovielens.py. It read input data, but give test error message about ”list index out of range”. Because there are no functionality in functions.

The `ast` module helps Python applications to process trees of the Python
abstract syntax grammar.

b) you can use “loadmovielens.py” and its function in “ASSG2_TMP.py” file by “reader”. Example use from file in “ASSG2_TMP.py” “reader.read_movie_lens_data()” read and return all database

function “print reader.give_me_movie_id('story', movie_dictionary)” show all movies with story, like Toy Story + its movie id number.
ToyStory id number is 1 and GoldenEye is 2

pseudo code: a = get all database values, where are values from story and Golden, sum their rating values,
b = then find toy or golden and sum their ratings values
then a/b in Jaccard_Coefficient

Kari Lerppinen

Harj 2 Introduction to machine learning 6.11.2015 st.id 0129186S6

Ex 3 b) we get distance between point x and y from Pythagorean definition

in $\mathbb{R}^n$ $d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$

we sum points from set one to
set two

to get distance from set one to set two, we can use center values from set one and two

Ex 3 c) we can count separate proximity values from data objects in set one and two and use average value from both sets

Ex 3 a) like in 3b, sum points from set and use center value or count minimum similarity of objects